Ji Yihong
1003040

1. ENTROPY

1.1 The Lagaragian problem for this optimization problem is

$$\mathcal{L}(\vec{p},\lambda) = -\sum_{i=1}^{n} p_i \log p_i + \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

Thus,
$$\frac{\partial \mathcal{L}}{\partial p_i} = -\left(\log p_i + p_i \cdot \frac{1}{p_i}\right) + \lambda$$
$$= -\log p_i - 1 + \lambda \quad , \forall i \in \{1,2,\ldots,n\}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{n} p_i - 1$$

Setting $\nabla \mathcal{L}(\vec{p},\lambda) = \vec{0}$ yields

$$-\log p_i - 1 + \lambda = 0 \ , \forall i \quad - ①$$
$$\sum_{i=1}^{n} p_i = 1 \quad - ②$$

From ① we know that

$$p_i = e^{\lambda - 1} \ , \forall i$$

and substitute into ②

$$n p_i = 1$$
$$\Rightarrow p_i = \frac{1}{n} , \quad \forall n$$

The Hessian matrix of $Ent(\vec{p}) = -\sum_{i=1}^{n} p_i \cdot \log p_i$ at $(\frac{1}{n}, \frac{1}{n}, \cdots \frac{1}{n})$ is

$$H = \begin{bmatrix} -\frac{1}{p_1} & & & \\ & -\frac{1}{p_2} & & \\ & & \ddots & \\ & & & -\frac{1}{p_n} \end{bmatrix}$$

, which is a diagonal matrix and with eigenvalue sitting on diagonal. Since the eigenvalue for the Hessian is all negative, the Hessian at $(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n})$ is negative definite and hence $ent(\vec{p})$ attains a local maximum here.

The uniform distribution maximize entropy function

1.2

We know that the lower bound for $ent(\vec{p})$ is 0, we only need to show the lower bound is attainable. Let exactly one of p_i be 1 and the rest of p_j be 0.

$$
\begin{aligned}
Ent(\vec{p}) &= -\sum_{i=1}^{n} p_i \cdot \log(p_i) \\
&= -(1) \cdot \log(1) + (n-1) \cdot \lim_{x \to 0} x \cdot \log x \\
&= 0 + (n-1) \cdot \lim_{x \to 0} \frac{\log x}{\frac{1}{x}} \\
&= (n-1) \cdot \lim_{x \to 0} (-x) \\
&= 0
\end{aligned}
$$

The constant random variable with degenerate distribution minimizes entropy.

2. SCHUR Complement

Consider

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}$$

Then

$$Ax + By = u \quad — ①$$

$$Cx + Dy = v \quad — ②$$

Assume D is invertible, from ② we have

$$Dy = v - Cx$$

$$y = D^{-1}(v - Cx)$$

Substitute into ① we have

$$Ax + BD^{-1}(v - Cx) = u$$

$$\Rightarrow Ax + BD^{-1}v - BD^{-1}Cx = u$$

$$\Rightarrow (A - BD^{-1}C)x = u - BD^{-1}v$$

Assume $A - BD^{-1}C$ is invertible,

$$x = (A - BD^{-1}C)^{-1}(u - BD^{-1}v)$$

Hence

$$x = (A - BD^{-1}C)^{-1}(u - BD^{-1}v)$$

$$y = D^{-1}(v - C(A - BD^{-1}C)^{-1}(u - BD^{-1}v))$$

By expansion,

$$x = (A - BD^{-1}C)^{-1}u - \underbrace{(A - BD^{-1}C)^{-1}}_{M}BD^{-1}v$$

$$y = -D^{-1}C\underbrace{(A - BD^{-1}C)^{-1}}_{M}u + \left(D^{-1} + D^{-1}C\underbrace{(A - BD^{-1}C)^{-1}}_{M}BD^{-1}\right)v$$

Let $M = (A - BD^{-1}C)^{-1}$, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Thus,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -M \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix}$$

where

$$M = (A - BD^{-1}C)^{-1}$$

Question3 Convolution Network

"C:\Users\JI YIHONG\Anaconda3\envs\6.86x\python.exe" "C:/Users/JI YIHONG/PycharmProjects/SML/main.py"

Test set accuracy: 12.48 %

Epoch: 1 ,loss: 2.3005735874176025

Epoch: 1 ,loss: 2.233872890472412

Epoch: 1 ,loss: 1.7953153848648071

Epoch: 1 ,loss: 1.2076956033706665

Epoch: 1 ,loss: 1.0014328956604004

Epoch: 1 ,loss: 0.9629236459732056

Epoch: 1 ,loss: 0.8627889752388

Epoch: 1 ,loss: 0.8257870674133301

Epoch: 1 ,loss: 0.7543894648551941

Epoch: 1 ,loss: 0.8580511212348938

Test set accuracy: 70.57 %

Epoch: 2 ,loss: 0.7097643613815308

Epoch: 2 ,loss: 0.768993079662323

Epoch: 2 ,loss: 0.530007004737854

Epoch: 2 ,loss: 0.7778584361076355

Epoch: 2 ,loss: 0.6932150721549988

Epoch: 2 ,loss: 0.7482984662055969

Epoch: 2 ,loss: 0.6694900393486023

Epoch: 2 ,loss: 0.6300978064537048

Epoch: 2 ,loss: 0.6084534525871277

Epoch: 2 ,loss: 0.6961829662322998

Test set accuracy: 75.87 %

Epoch: 3 ,loss: 0.5677799582481384

Epoch: 3 ,loss: 0.625316858291626

Epoch: 3 ,loss: 0.42085927724838257

Epoch: 3 ,loss: 0.6807870268821716

Epoch: 3 ,loss: 0.5829302072525024

Epoch: 3 ,loss: 0.6529091596603394

Epoch: 3 ,loss: 0.576701283454895

Epoch: 3 ,loss: 0.545748770236969

Epoch: 3 ,loss: 0.5735911726951599

Epoch: 3 ,loss: 0.5683251023292542

Test set accuracy: 79.1 %

Epoch: 4 ,loss: 0.5002447366714478

Epoch: 4 ,loss: 0.5686485171318054

Epoch: 4 ,loss: 0.3665580153465271

Epoch: 4 ,loss: 0.6052994728088379

Epoch: 4 ,loss: 0.5413296818733215

Epoch: 4 ,loss: 0.5810123682022095

Epoch: 4 ,loss: 0.508638858795166

Epoch: 4 ,loss: 0.5155031085014343

Epoch: 4 ,loss: 0.5595162510871887

Epoch: 4 ,loss: 0.4903385043144226

Test set accuracy: 80.73 %

Epoch: 5 ,loss: 0.461159348487854

Epoch: 5 ,loss: 0.5271494388580322

Epoch: 5 ,loss: 0.3390577733516693

Epoch: 5 ,loss: 0.5505642890930176

Epoch: 5 ,loss: 0.5128136873245239

Epoch: 5 ,loss: 0.5266345143318176

Epoch: 5 ,loss: 0.46538498997688293

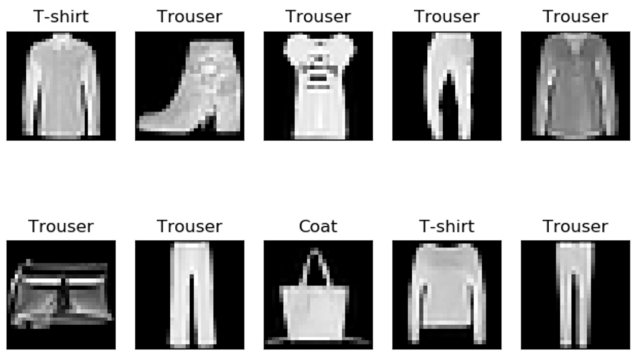
Epoch: 5 ,loss: 0.5168305039405823

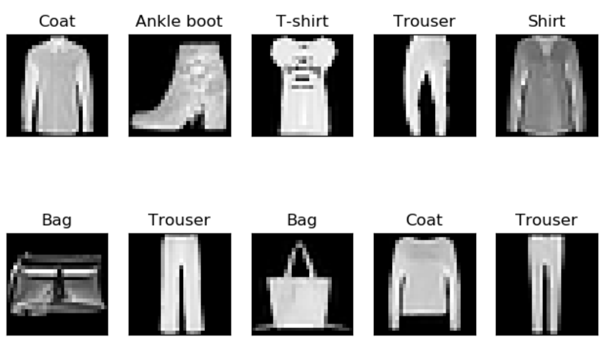
Epoch: 5 ,loss: 0.5486657023429871

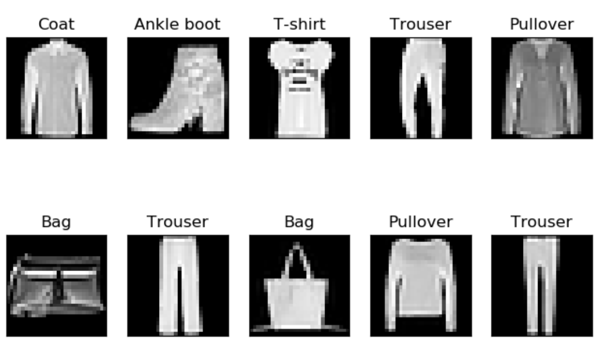
Epoch: 5 ,loss: 0.43864670395851135

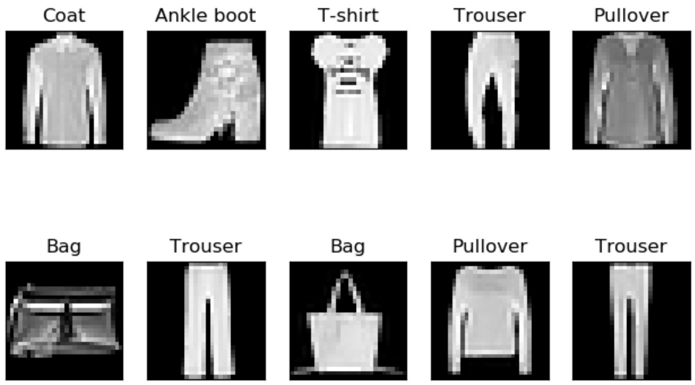
Test set accuracy: 82.05 %

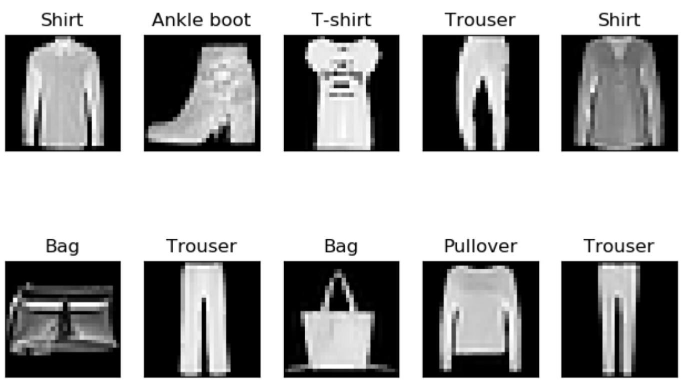
Process finished with exit code 0

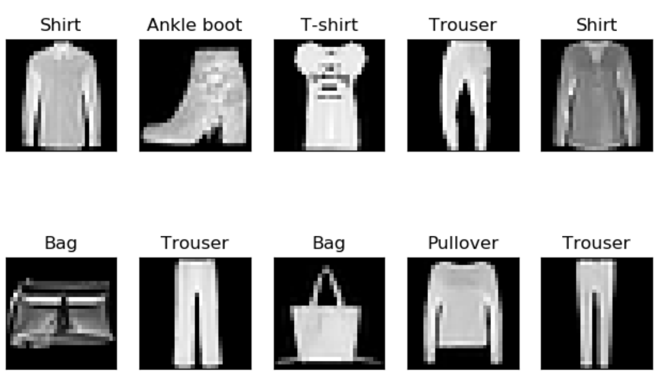






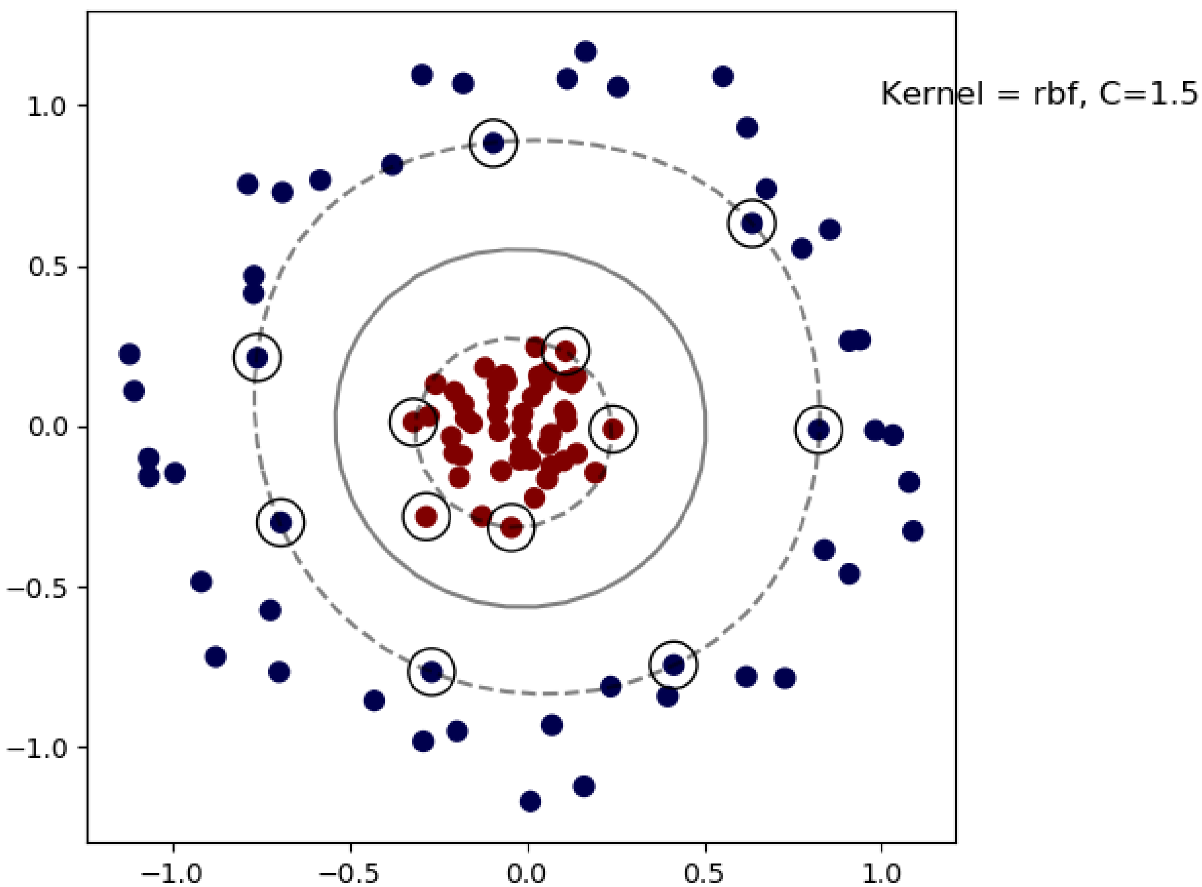

Coat
Ankle boot
T-shirt
Trouser
Pullover
Bag
Trouser
Bag
Pullover
Trouser


Shirt
Ankle boot
T-shirt
Trouser
Shirt
Bag
Trouser
Bag
Pullover
Trouser


Shirt
Ankle boot
T-shirt
Trouser
Shirt
Bag
Trouser
Bag
Pullover
Trouser

Question 4 Support Vector Machines

4.1 Kernel = rbf, C=1.5



4.2 kernel =linear ,C=1 for left plot

kernel=linear C=0.3 for right plot